1. In biodiesel study, methyl ester was produced from waste canola oil. In experiments, it was measured what kind of effect the factors $X_1 =$ Time (15,30,45min), $X_2 =$ Temperature (240,255,270C), and level of Methanol/Oil weight ratio (1,1.5,2), $X_3 =$ Methanol, have on yield of methyl ester, $Y =$ Yield. Data obtained from experiments is available in a file canoladiesel.txt.

	Time	Temp	Methanol	Yield
1	15	240	1.0	1.5
2	15	240	1.5	3.2
3	15	240	2.0	3.8
4	15	255	1.0	2.2
5	15	270	1.0	8.9
6	15	270	2.0	13.6
7	30	240	1.0	1.5
8	30	240	2.0	12.3
9	30	255	1.5	11.4
10	30	255	1.5	13.6
11	30	255	1.5	12.7
12	30	255	2.0	18.5
13	30	270	1.5	60.9
14	45	240	1.0	4.4
15	45	240	1.5	16.5
16	45	240	2.0	24.5
17	45	255	1.5	62.8
18	45	270	1.0	96.4
19	45	270	2.0	102.0

Source: S. Lee, D. Posarac, N. Ellis (2012). "An Experimental Investigation of Biodiesel Synthesis from Waste Canola Oil Using Supercritical Methanol," Fuel, Vol. 91, pp. 229-237.

(a) Let us assume $Y_i \sim N(\mu_i, \sigma^2)$. Consider modeling the expected value μ_i of the response variable $Y = \mathsf{Yield}$ by only using $X_1 = \mathsf{Time}$ as an explanatory variable with the model

$$\mathcal{M}_{1_{\mathsf{inverse}}}: \quad \frac{1}{\mu_i} = \beta_0 + \beta_1 x_{i1}.$$

Calculate the maximum likelihood estimate for the expected value μ_{i_*} when $x_{i_*1} = 40$. (1 point)

(b) Let us assume $Y_i \sim Gamma(\mu_i, \phi)$. Consider the models

$$\begin{aligned}
\mathcal{M}_{1_{\mathsf{identity}}}: &\quad \mu_i = \beta_0 + \beta_1 x_{i1}, \\
\mathcal{M}_{1_{\mathsf{inverse}}}: &\quad \frac{1}{\mu_i} = \beta_0 + \beta_1 x_{i1}, \\
\mathcal{M}_{1_{\log}}: &\quad \log(\mu_i) = \beta_0 + \beta_1 x_{i1},
\end{aligned}$$

Which model fits the best to the data if the choice of model is done based on the AIC value? (1 point)

i. $\mathcal{M}_{1_{\mathsf{identity}}}$,
ii. $\mathcal{M}_{1_{\mathsf{inverse}}}$,
iii. $\mathcal{M}_{1_{\log}}$.

(c) Let us assume $Y_i \sim Gamma(\mu_i, \phi)$. Consider modeling the expected value μ_i of the response variable $Y = \mathsf{Yield}$ by the following model

$$\mathcal{M}_{1|2|3_{\log}}: \quad \log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}.$$

Calculate the maximum likelihood estimate for the expected value μ_{i_*} when $x_{i_*1} = 40$, $x_{i_*2} = 260$, and $x_{i_*3} = 1.75$. Calculate also the 95% confidence interval. (2 points)

(d) Let us assume $Y_i \sim Gamma(\mu_i, \phi)$. Test at 5 % significance level, is the explanatory variable X_3 statistically significant variable in the model

$$\mathcal{M}_{1|2|3_{\log}}: \quad \log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}.$$

Calculate the value of the test statistic. (1 point)

(e) Consider the model

$$\mathcal{M}_{1|2|3_{\log}}: \quad \log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3},$$

and competing distributions $Y_i \sim N(\mu_i, \sigma^2)$, $Y_i \sim Gamma(\mu_i, \phi)$, and $Y_i \sim IG(\mu_i, \phi)$. Study under different distributional assumptions how Pearson's residuals are behaving. Based on your analysis, under which distribution the model fits the best to the data? (1 point)

i. $Y_i \sim N(\mu_i, \sigma^2)$,
ii. $Y_i \sim Gamma(\mu_i, \phi)$,
iii. $Y_i \sim IG(\mu_i, \phi)$.

2. Consider the following data set ratstime.txt:

```
  time poison treat
1 0.31      I     A
2 0.82      I     B
3 0.43      I     C
4 0.45      I     D
5 0.45      I     A
6 1.10      I     B
.
.

Effect of toxic agents on rats

Description

An experiment was conducted as part of an investigation
to combat the effects of certain toxic agents.

A data frame with 48 observations on the following 3 variables.

time
survival time in tens of hours

poison
the poison type - a factor with levels I II III

treat
the treatment - a factor with levels A B C D
```

The response variable is $Y =$ time and the explanatory variables are X_1=poison and X_2=treat.

(a) Model the data with the main effect model

$$\mu_{jh} = \beta_0 + \beta_j + \alpha_h.$$

Distributional assumption could be either $Y_i \sim N(\mu_{jh}, \sigma^2)$, $Y_i \sim Gamma(\mu_{jh}, \phi)$, or $Y_i \sim IG(\mu_{jh}, \phi)$. Based on your analysis, which one is most suitable in this case?

(1 point)

(b) Regardless what was your solution to the question (a), assume $Y_i \sim Gamma(\mu_{jh}, \phi)$. Based on the mean square error value $\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_{jh})^2}{n}$, which link function fits best to the data:

$$\begin{aligned} \mu_{jh} &= \beta_0 + \beta_j + \alpha_h, \\ \log(\mu_{jh}) &= \beta_0 + \beta_j + \alpha_h, \\ \frac{1}{\mu_{jh}} &= \beta_0 + \beta_j + \alpha_h? \end{aligned}$$

(1 point)

(c) Assume $Y_i \sim IG(\mu_{jh}, \phi)$. Consider the hypothesis

$$\begin{aligned} H_0 &: \log(\mu_{jh}) = \beta_0 + \beta_j + \alpha_h, \\ H_1 &: \log(\mu_{jh}) = \beta_0 + \beta_j + \alpha_h + \gamma_{jh}. \end{aligned}$$

Select appropriate test statistic to test the above hypotheses. Calculate the value of the test statistic.

(1 point)

(d) Assume $Y_i \sim Gamma(\mu_{jh}, \phi)$. Consider the model

$$\log(\mu_{jh}) = \beta_0 + \beta_j + \alpha_h.$$

Create 80 % prediction interval for new observation y_f, when

$$x_{f1} = II, \qquad x_{f2} = B$$

(2 points)

(e) Assume $Y_i \sim Gamma(\mu_{jh}, \phi)$. Consider the model

$$\log(\mu_{jh}) = \beta_0 + \beta_j + \alpha_h.$$

Construct 80% prediction interval for the (predictive) effect size difference $y_{2f} - y_{1f}$ when explanatory variables are changed from the values

$$x_{1f1} = I, \qquad x_{1f2} = D$$

to the values

$$x_{2f1} = II, \qquad x_{2f2} = B.$$

(1 point)

3. (a) In case of generalized linear model $g(\mu_i) = \beta_0 + \beta_1 x_i$, find the inverse function g^{-1} (that is, solve what form the expected value μ_i has), when the link function g is

   i. $\sqrt{\mu_i} = \beta_0 + \beta_1 x_i$,
   ii. $\frac{1}{\mu_i^2} = \beta_0 + \beta_1 x_i$,
   iii. $\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_i$.

   (2 points)

   (b) Let us assume $Y_i \sim IG(\mu_i, \phi)$. Consider the model

   $$\log(\mu_i) = \beta_0 + \beta_1 \log(x_i).$$

   Let the estimates of the parameters β_0, β_1, ϕ be as $\hat{\beta}_0 = 1, \hat{\beta}_1 = 0.5, \tilde{\phi} = 0.05$.

   i. Calculate the maximum likelihood estimate for the expected value μ_i when $x_i = 5$.
   ii. Calculate the Pearson residual

   $$o_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\mathrm{Var}}(Y_i)}},$$

   when $x_{i_*} = 5$ and the observed value is $y_i = 12$.

   (2 points)

   (c) Consider the normal distribution with the identity link function

   $$\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}),$$
   $$\boldsymbol{\mu} = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}.$$

   What kind of more simplified form the likelihood equations

   $$\frac{\partial l(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta}} = \mathbf{u} = \mathbf{X}'\mathbf{D}\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$$

   have in this case? Note that

   $$\mathbf{D} = \begin{pmatrix} \frac{\partial \mu_1}{\partial \eta_1} & 0 & \dots & 0 \\ 0 & \frac{\partial \mu_2}{\partial \eta_2} & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \frac{\partial \mu_n}{\partial \eta_n} \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} \mathrm{Var}(Y_1) & 0 & \dots & 0 \\ 0 & \mathrm{Var}(Y_2) & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \mathrm{Var}(Y_n) \end{pmatrix}.$$

   Can you obtain the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ directly by solving likelihood equations with respect to $\boldsymbol{\beta}$?

   (2 points)